DOCUMENT RESUME

ED 385 324 JC 950 398

AUTHOR Armstrong, William B.
TITLE Validating Placement Tests in the Community College: The Role of Test Scores, Biographical Data, and Grading Variation.
PUB DATE 95
NOTE 35p.; Paper presented at the Annual Forum of the Association for Institutional Research (35th, Boston, MA, May 28-31, 1995).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Access to Education; Community Colleges; Educational Background; Ethnic Groups; *Grading; Multiple Regression Analysis; *Predictive Validity; Predictor Variables; *Standardized Tests; *Student Placement; *Test Validity; Two Year Colleges

ABSTRACT

Concerns over equity, access, and fairness of placement testing have resulted in state regulations requiring validity evidence of the placement test with respect to a criterion variable, such as final course grades. A study was conducted to assess the validity of placement tests used to place students in pre-collegiate and college-level English and mathematics courses. The study examined the correlation between scores received on placement tests and student grades in English and mathematics courses. Also examined were other variables which helped to explain the variance in course grades. To determine which variables to include in the regression equation, several cross-tabulations were conducted to examine the association between ethnic/racial groupings; educational background; commitment to college attendance; and instructor grading variables. The study indicated that any model that used final course grade as the criterion for the validity of a placement rule would likely fail to account for a significant source of variation if instructor grading variability was not included. This finding has implications for validation policy and instructional issues across the curriculum. The use of instructor codes in test validation research is politically sensitive; however, it appears that in the interest of student equity and access, these critical instructional issues should be discussed. (KP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 385 324

Validating Placement Tests in the Community College: The Role of Test Scores, Biographical Data, and Grading Variation

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
W. Armstrong
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the Association for Institutional Research
35th Annual Forum

William B. Armstrong
Director, Insitutional Research
San Diego Community College District

JC 950 318

BEST COPY AVAILABLE

Abstract

Concerns over the equity, access, and fairness implications of placement testing resulted in state regulations requiring that community colleges provide validity evidence of the placement test to a criterion variable such as the final grade in the course. Preliminary analyses suggest several biographical variables as highly related to course success. Analyses suggested however that including the instructor in the model explained the largest proportion of the variance in final grade. Implications of these findings for test validation studies and Point-to-Point theory will be discussed to inform test policy and practice in a variety of educational settings.

News Item:

A lawsuit filed by the Mexican-American Legal Defense Fund (MALDEF) alleged that Fullerton College used scores on a placement test to force students to take non-credit, remedial courses before they could enroll in courses carrying college credits that were transferable to four-year colleges and universities. The suit, which included 19 other colleges charges that these colleges have violated the Matriculation Law and due-process guarantees of the California constitution by refusing to allow some students to enroll in college-level courses for credit.

(Chronicle of Higher Education, June 1, 1988)

Introduction and Background

The Seymour-Campbell Matriculation Act of 1986 (AB-3, 1986) mandated that California community colleges design and implement a Matriculation program to improve student outcomes. This reform was enacted in response to concerns raised by college constituents and legislators about declining academic standards and student achievement (Alkin, 1991). This reform was intended to improve student outcomes in the two-year colleges by more clearly delineating the responsibilities of students as participants in the educational process and providing enhanced support services to help students meet those responsibilities. The matriculation program includes five components that directly affects students. One of these mandated components is an assessment and placement component that is heavily reliant on the use of standardized tests for student placement in different levels of English and mathematics. In

many cases, scores on placements tests are used as skill level pre-requisites for courses in other departments as well. It is the assessment and placement component of matriculation that this study focuses on. Specifically, what is the relation between scores achieved on placement tests and final grade in selected courses? And, what other factors must be considered when institutions attempt to sort and place students at various levels (pre-collegiate and college-level) in the curriculum in ways that maximize their chances for success? Additionally, this study should also provide some guidance to other researchers and evaluators seeking to validate the use of placement tests, particularly when the criterion for test validation is the final grade achieved in the course of interest.

Matriculation and the Growth of Placement Testing

The implementation of matriculation in the California community colleges helped to fuel rapid growth in the use of standardized and locally developed tests for placing students into different levels of coursework. A survey conducted by the State Chancellor's Office of the California Community Colleges (SCOCCC, 1989) indicated that since the passage of the Matriculation legislation, almost 80% of the community colleges in California now use standardized tests to assess and make determinations about student aptitude for English and mathematics course placement. Although colleges are by no means uniform in local applications, adaptation, and use of these tests in placing students, nonetheless, growth in skills testing appears to be among the most prominent outcomes of the matriculation reform.

Equity and Access Implications

The MALDEF lawsuit cited in the opening paragraph adds another dimension to an on-going testing debate within higher education and the education community in general (Astin, 1991; The National Commission on Testing and Public Policy, 1990). The access and equity issues raised by testing critics underscores the importance of examining the validity of standardized tests in the allocation of educational opportunity. This issue is particularly

important in light of the perceived conflict between the open door mission of the community colleges, the differential test performance of certain cultural and linguistic minority students, and recent state mandates in testing, placement, and test validation requirements (State Chancellor's Office, 1992).

Test Validation and the Classroom Door

A closely related issue involving the validation of testing as it relates to course performance is the evaluation of student performance by instructors. Concerns over equity and access as they relate to the predictive validity of a placement test must necessarily involve the evaluation of student course performance by individual instructors. Prior research suggests that there is often tremendous variation in how performance is judged by individuals (Sticht, 1986, & Sticht, 1975; Cronbach, 1965). The application of "Point-to-Point" theory as described by Asher (1974) suggests that the use of cut scores on a placement test to predict the often normative outcome variable of final grade on a five-point scale will generally result in unreliable coefficients. What is deemed as excellent performance by one may be seen by another as average. Or there may be departmental or institutional policies regarding how high or low ratings of performance are allocated. Thus if a test fails to predict a grade, we may not know if this is due to the test, or the instability of the criterion variable of final grade. If grades or ratings of performance are considered one of the criteria by which placement tests are judged in order to address issues of access and equity, the variation inherent in the judgment process must also be considered. In higher education, the evaluation of performance and the assignment of grades is strictly within the purview of the instructor. However, to address equity and access concerns, state law mandates that the colleges judge the validity of placement tests by how well they predict student course grades. Thus does test validation policy edge closer to the classroom door.

Equity and access concerns also militate strongly for the development of alternate models that include other cognitive and affective factors that demonstrate a relationship with student course success to more fairly and accurately place students. The use of multiple

measures to place students is also based in state policy. Rather than relying on a single indicator of aptitude to select students for different levels of instruction, colleges are directed to assess and include other relevant data known to be related to student success (State Chancellor's Office, 1992). These data might include GPA in high school, prior course work, educational goal, the interaction of the number of units planned and employment hours the student plans to work, etc. This information should result in a more accurate and equitable judgment of what level to place the student in the curriculum. Such a multiple measures model would meet state mandates for access by providing alternate indicators of achievement, motivation, and aptitude often not measured on a test. This model could then be used to improve the placement system and influence policy favorably for student success. It is these two areas, test validation and the development of a student course success model, that this study will address.

Purpose of this Study

There are two primary questions addressed by this research study. One question is concerned with examining the validity of placement tests used to place students in pre-collegiate and college level English and mathematics coursework. Both Title V and Federal Civil Rights legislation require that institutions validate the use of assessment tests in the placement and referral of students. In fact, in addition to a lengthy list of regulations governing the use of tests and their effects, the State Chancellor's office of the California Community Colleges has mandated that in order for a test to be considered a valid predictor of course success, colleges must demonstrate a **.35** correlation between the test score and the criterion measure of course grade (State Chancellors Office, 1992).

Research Question One: Relation of Placement Test Score to Course Grade

The first question attempts to address this policy issue by reporting the observed correlation between scores received on the Assessment and Placement Services for the Community College (The College Board, 1984), in English and Reading, and scores received by

students on the Mathematics Testing Diagnostic Project (MTDP) tests and the grades these students received in English and mathematics courses. Question one specifically asks:

What is the statistical validity of test instruments and placement recommendations used by the San Diego Community College District (SDCCD) to place students in pre-collegiate and college level English and Mathematics courses?

Approach

Validity will be assessed using correlational methods to assess the relationship between placement scores and student performance as defined by course grade converted into a numeric format (A=4, B=3, C=2, D=1, F=0, with W and Drop coded as missing data). Additionally, 2 X 2 classification tables were used to assess the degree of association between the cut scores used to place students and successful completion of the course as defined by attaining a 'C' grade or better.

Research Question Two: Explaining Variance in Course Grades

The second question may be viewed as an outcome of the first. A review of the literature examining the relation between placement decisions based on standardized tests and performance ratings suggests that the correlation coefficients obtained between paper and pencil tests and individual ratings of performance by instructors, employers, and supervisors is often very low (Cronbach & Gleser, 1965). In fact, in one study conducted for the military services, the resulting coefficients between the predictor variables of aptitude test scores (AFQT) and the dependent variable of supervisor ratings were so low, that the researchers decided to focus solely on the relationship between the aptitude tests and job knowledge tests (a paper and pencil test) and eliminated supervisor's rating from their analysis (Sticht, 1975). This review of prior research suggested that the analysis planned to answer question one would yield low correlation coefficients and that other factors such as prior performance, commitment to college, outside responsibilities, and instructor grading variability must be included in order to understand what

might explain a weak relationship between test scores and final grade. This led to formulation of research question two:

What combination of variables may help to explain the variance found in grades (A - F) for students in pre-collegiate and college level English and Mathematics courses?

This question is intended to address the equity and policy considerations of student assessment and placement in the community colleges. This will be accomplished by identifying factors for consideration when state mandates for placement test implementation, test validation criteria (i.e.; the .35 correlation mentioned earlier), student access, and fairness are made operational in the college setting.

Approach

To address the second research question, data from a counselor survey containing several variables related to student input, commitment, educational background, and environmental variables were merged with the SDCCD student information database(see Table 1). Input variables include demographic data such as income, age, race, veterans status, native language, and sex of responcent. Educational background data include such data as years out of school, high school GPA, high school of attendance, and highest level of English and mathematics courses completed while in high school. Environmental variables available for use in this study include campus of attendance, day or evening attendance, college major, and instructor code to assess grading variability. What I have termed here as "commitment" variables include the importance of college to the student and to those closest to the student, employment hours planned while attending classes, educational goal, the number of units planned for this and the next term. Again, the dependent variable is final course grade in the English or mathematics courses enrolled in by students in the database. Table 1 summarizes the variables, data sources, and analyses for both of the primary research questions.

TABLE 1

RESEARCH QUESTIONS,

TYPE AND SOURCE OF VARIABLES,

AND ANALYSES

Question	Variables	Source	Analysis
Assessment Test Validation	**Independent:** APS Reading and English Placement Test Scores MDTP Mathematics Placement Scores **Dependent:** Course Grades (A=4, B=3, C=2, D=1,0=F)	SDCCD Student Database and Instructor Grade Rosters	Correlation and Crosstabulation
What Combination of Variables may help to Explain the Variance in Grades for Selected Pre-collegiate and College level courses?	**Independent:** ***Input*** : Ethnicity, Gender, Age, Disability, Test scores **Independent:*Education*:** High school GPA, Years out of school, Grade in last English or Math class. High School Education., Income, High School attended, Years of high school English **Independent:*Commitment*:** Hours worked per week, Importance of college to student *and* to people closest to student, Educational goal, Units planned for this and next term How definite choice of major, Need for educational services. **Independent:** ***Environmental***: College of attendance, Instructor ID **Dependent**-Final course grade	SDCCD Student Test Score and Demographic Databases Student Surveys merged with SDCCD Student Database. Instructor Grade Rosters, and Drop/ Withdraw Records	Crosstabulation and Multiple Regression

There were also policy and equity considerations behind the formulation of the multiple measures approach to explaining course performance. Although AB-3 mandates the use of testing in the community colleges for placement purposes, state level policymakers and special interest groups such as MALDEF were also wary of the use of testing to screen, sort, and sift students according to a single "cut" score. As a result, the state chancellor's office was directed to ensure that colleges not rely on a single measure or cut score to sort students according to ability level for placement purposes (State Chancellor's Office, April, 1992, p. 6, section 55521). To be in compliance with state and federal civil rights laws and to respect student rights, colleges were mandated to use multiple measures of ability in determining placement. These multiple measures might include high school GPA, number of hours worked per week, age, educational goal, transcript review, perceived level of commitment, motivation, language fluency, years out of school, etc., in conjunction with test scores. These several measures would be used to yield a composite picture of the matriculant, and in the judgment of the professional conducting the assessment (usually a counselor), an appropriate placement would be made. Thus test scores were to be part of an overall assessment of individual aptitude, not the sole criterion of ability to benefit from instruction at a certain level. The need for such a model then is rooted in student equity and access, as well as educational policy, regardless of the relation between test scores and course performance.

Although the criterion variable of final course grade would be the same for both research questions, the independent or explanatory variables would include not only placement test scores, but other input, educational background, affective, and environmental data that might help to explain variance in final course grades in pre-collegiate and college level English and mathematics courses. Also this model departs somewhat from other models reviewed that attempted to predict student success in specific community college courses (Hughes & Nelson, (1991); Gillespie (1993)) in that the instructor assigning the grade was also included in the model explaining variance in course grades.

Background of the Matriculation Reform

As the traditional open-access institutions for a majority of historically under-represented groups, the community colleges have witnessed tremendous growth over the past ten years in the demand for basic skills courses in English, Mathematics and English language instruction. With evidence of declining academic skills, primarily in the basic skills of reading, writing, and mathematics, educators at the post-secondary level have witnessed a sharp rise in the number of students requiring basic skills instruction and remediation (Cohen, 1982). These observations, combined with a decline in student transfer rates, graduation, and persistence rates, led to a thorough examination of the mission and purposes of the two-year colleges in California.

The Matriculation Act was passed in 1986 after a flurry of national and state reports had been released describing a steady decline in academic standards and performance of the two-year colleges and education in general. Paralleling similar national reports such as "A Nation at Risk," studies conducted by the Academic Senate for the California Community Colleges suggested that academic standards were declining and, with them, the quality of educational programs. In response to that concern, the Academic Senate adopted a resolution in the fall of 1982 entitled, "Matriculated Student." This resolution stated that community college students seeking a degree or certificate should be designated as "matriculated," and be held to certain requirements in order to maintain that status.

In response to this resolution, the State Chancellor appointed a Task Force on Academic Quality to propose a model process that might be used to help community college students define their educational goals and make informed educational choices to achieve them. The concept of the "matriculating" student was a core element of matriculation program development. Following a pilot test of the program in sixteen community colleges in 1984, the Board of Governors of the California Community Colleges adopted a plan entitled "Student Matriculation: A Plan for Implementation in the California Community Colleges (June, 1984).

Following the review of the state Master Plan for Higher Education in 1986 and the subsequent endorsement of the matriculation concept by the master plan review committee, the

Seymour-Campbell Matriculation Act was passed in 1986, but because of uncertain state revenues that year, not funded until 1988.

Matriculation Purpose and Goals

A primary purpose of the Matriculation legislation was to make the educational process for students at the colleges more intensive. This was to be accomplished by compelling students to define their educational goals early in their program of study and make "realistic" educational choices as to coursework, and short- and long-term educational and occupational goals (State Chancellor's Office, 1989). This reform mandates that the colleges develop and implement a matriculation program with seven distinct components, five of which directly affect students. These components include an **admissions** program that guarantees access, **orientation** to college services, pre-enrollment **skills assessment and placement** advisement for course selection, **counseling** for educational planning and referral to support services, and **follow-up** on student progress.

Matriculation Reform and Placement Testing: A Response to Diversity?

The community colleges serve a diverse population. Learners coming to the colleges do so for a variety of reasons and motivations. As the purposes of students have diversified, so too have the students themselves. Increasingly community college students are older, attending part-time, female and more likely to be non-white than ever before. This national trend is particularly evident here in California where growth in non-traditional, non-white enrollments has proceeded apace since 1980. Accommodation of these "new majority" students is being accomplished in the college, but not without problems. As more students come to the colleges with increasingly divergent backgrounds, the colleges are having to deal with diversity in a magnitude greater than many had anticipated.

The use of standardized tests may be seen as an outcome of the increasing diversification of the community college student population. The concerns raised by the academic senate

regarding the need to group students according to ability level and place them in courses where they would be most likely to benefit from instruction appear to have been addressed through mass implementation of standardized testing for placement purposes. This despite concerns raised by some legislators, educators, and community groups that the use of these tests are in conflict with the open access instructional mission of the community colleges. This also despite a lack of evidence that placement tests would accurately predict a student's performance in class, or the grade received.

State Matriculation Regulations Regarding Testing Practices and Use

State legislative concern over the use of tests to screen, sift, and sort individuals for education, training, and employment is not new. Following a series of public discussions, policy papers, and legislative hearings, the legislature passed the first "truth in testing" law in California in 1978 (Senate Bill 2005, Dunlap). This "test disclosure" legislation required developers of standardized tests to file financial disclosure statements, copies of test materials, and answers to test questions with appropriate state agencies. Interestingly, many of the current arguments regarding testing in the 1990's can be found in the testimony and hearings taking place in the 1970's.

Passage and subsequent implementation of the matriculation act renewed the legislative and public focus on the use of placement tests in the community colleges. Among proponents of open access and expanding opportunities in the community colleges, there are few issues that generate as much heat as testing and assessment. The basis for much of this disagreement in the applications of testing is not difficult to understand, given that certain traditionally under-represented students such as poor, black, and Latino students tend to achieve lower scores on tests of educational achievement and aptitude and that most universities and colleges rely heavily on scores on standardized tests and high school grade point averages for admissions or for entry into high demand occupational programs.

The continuing over-reliance of such measures by college and university personnel will make it very difficult for any educationally underprepared group to be represented equally or

proportionately in higher education (Astin, 1991). Although the community colleges in California do not use standardized tests as an admissions screen, tests are used to select applicants to certain high demand training and occupational programs (ACT/AACJC, 1985). Critics of the use of standardized tests in these institutions argue that the very *use* of these tests constitute a psychological barrier for some groups.

In August 1992, the state chancellor's office of the California Community Colleges published a document entitled, *Standards, Policies, and Procedures for the Evaluation of Assessment Instruments Used in the California Community Colleges* (State Chancellor' s Office, 1992). This document sets forth criteria to be met by colleges and districts when they use assessment tests for placement purposes. All colleges must address the following standards when evaluating their use of assessment instruments:

Minimize or Eliminate Cultural Bias
Monitor Disproportionate Impact on Particular Groups of Students by Age, Ethnicity, Sex, or Disability Grouping
Establish Local Norms
Establish the Local Reliability and Validity of Placement Tests
Validity of "Cut Scores" used for Placement
The Use of "Multiple Measures" in Course Placement
The Predictive Validity of the Placement Test to the Outcome Measure (i.e., demonstrate a .35 Correlation)

Evidence must be gathered to address each of these specific regulations for review by visiting audit teams. For example, for the test validation standard, the regulations state that for colleges using a locally developed test, or a modified commercially developed test, "the correlation between the test and a student's readiness to assimilate course content (e.g., mid-term grade, instructor or student evaluation, end of course grade) ; ...should be greater than or equal to .35" (p. 25). For institutions that use an unmodified test for placement, they must demonstrate

that students scoring above the cut score have a "relatively greater expectancy of success" (obtaining a "C' grade or higher) than students scoring below the score or score range (p. 21).

It is through the development and enforcement of these standards that the state interest in protecting student access and minimizing the use of tests as an obstacle to enrollment is supposed to be safeguarded. Through the promulgation of test validation mandates, state officials have sought to ensure that scores achieved on placement tests are at least moderately related to the criterion of grade received in the course for which the test was designed to place students (i.e., English placement test to English course grade, mathematics test to mathematics course grade, etc.). Additionally, they have mandated the use of other measures in the placement decision so that students are not placed on the basis of a single test score. These policies are the grist from which the original research questions were developed. However, the findings obtained from this study suggest that state policymakers and staff did not consider a major source of variation in the criterion variable of course grade, that of instructor grading variation.

Method

To answer both research questions, a database was developed that included the variables displayed in Table 1. A request was submitted to the SDCCD data processing unit for the creation of a ASCII file that contained several variables relevant to the study, following as closely as possible the recommendations put forth by Astin (1991) for database construction. Because of recent database conversions and subsequent problems with the 1991-1992 student data system, the data-processing staff recommended that the first-time entering student cohort for fall, 1990 be used for this study.

Students included in the database were restricted to those first time students who had taken at least one of the assessment tests (reading, writing, or mathematics) and had enrolled that same fall semester in an English or mathematics course. To test the validity of the placement tests across all levels of course difficulty, course data included English and mathematics courses across at least three levels of difficulty (2 levels below college, 1 level below transfer level

(typically offered only for credit toward the associates degree), and transfer (UC/CSU) level), the database was organized with course as the grouping variable. For the database as a whole, a student record would appear for every class they enrolled in. For example a student enrolling in English 101, History 1A, and PE would appear in the database three times. Analyzing the data by course number minimizes the likelihood that a student would appear more than once in a given analysis. For example, it was unlikely that a student would enroll in two sections of English 51, or two sections of college algebra. Examination of the database revealed that this could occur if a student enrolled in more than one section of a course and failed to drop one of the courses before the withdrawal deadline. This happened less than 3% of the time however. In this case a 'W' would appear on the duplicate record. However, this problem was eliminated by dropping the duplicate record from the analysis when the student received a 'W' in the duplicated course.

The Counselor Survey

The SDCCD counseling staff has for the last three years administered a questionnaire to students asking them several questions related to their educational plans, aspirations, background, and commitment to college attendance (see Table 1 for a summary of the data elements). Originally administered to aid counselors in the placement process and to gather multiple measures of student aptitude, the data were left essentially unanalyzed until this study. Since 1988, thousands of these surveys have been completed and filed away with the student records in the registrars office. This "fugitive" database of student survey responses represented a potentially rich source of information regarding the factors predicting student success and commitment to college particularly when analyzed together with the existing student database. Data from this survey were hand-entered into a Dbase III file, and using the student ID as the matching variable, merged with the student enrollment and course performance data using the JOIN procedure in SPSS-PC+. After securing permission from the dean of student services at each of the colleges, a codebook was created and a database structure prepared. Data entry was started late in the fall, 1992 quarter and continued throughout the winter, 1993 quarter. In all

6,570 surveys were entered. These were merged into a single file in SPSS and saved as a system file. Once merged, there were 3,569 student records for the English course analysis, and 3,011 records for the mathematics course analysis.

Database Problems and Constraints

Although the counselor surveys were generally complete, there were much missing data for students whose did not identify English as their native language. It appears that when students were unable to read or understand a question, they generally skipped it. Unfortunately this seems to have occurred most often with the Asian and Latino respondents, two of the fastest growing cohorts of the community college student population in California and the SDCCD. In future analyses using this database, the feasibility of using substituting the mean for missing data will be explored.

Another problem with the counselor survey is the questions that weren't included. Although some potentially valuable questions were included, other questions that have long been identified in the literature and found on other surveys such as the CIRP or the Community College Student Experiences Questionnaire (CCSEQ) were not. This may have resulted in a mis-specified or at least an under-specified model for research question two.

The dependent variable of final grade also presented some difficulties. Although not the only nor error-free outcome measure, student final grades have often been chosen as the criterion variable in research studies measuring student performance. This is probably related to ease in collecting and reporting. However, because the use of final grades as a criterion measure for placement testing is rooted in state policy and instructional practice, the extent to which grading practices vary from instructor to instructor needs to be identified as to the impact it has on the statistical data generated for research purposes, as well as concerns of student equity.

Another problem is the 'W' grade (Withdrawal). This grade notation appeared in approximately 20 percent of all student grades. The W grade is troublesome because of the uncertainty about what it represents. In some instances, the W could be considered a non-

successful grade if students who are performing below a certain level withdraw from the course to avoid a low grade. On the other hand, a student may be doing well in a course and drop late in the term for reasons unrelated to academic ability or commitment to college attendance. Another source of variation in the assignment of the W might be the instructor or policy of the respective departments. An instructor might assign a W to a student who stops attending without formally dropping the course rather than assigning a failing grade. In the analysis conducted to address question one, correlations were computed between test score and final grade without W's included and with W's included.[1] When W's were included, the resulting correlations were lower than when W's were excluded. In subsequent analyses for this study, the decision was made to not include the W's in both the test validation analysis and the grade variance analysis.

Analysis

The English, reading, and mathematics test scores were correlated with the final grade obtained in three levels of English course, and Math 54 (Elementary Algebra) using the Correlation procedure in SPSSPC+. Only one level of mathematics course was selected to be used in this study because after listwise deletion of missing data, other courses had insufficient cases for analysis. The results of correlational analysis are shown in Table 2 below.

[1] When W's were included, the coding was changed to A=5, B=4, C=3, D=2, F=1, and W=0.

TABLE 2

CORRELATION OF APS READING AND ENGLISH TEST SCORES WITH GPA BY LEVEL OF ENGLISH AND CORRELATION OF MTDP MATHEMATICS TEST WITH GPA FOR ELEMENTARY ALGEBRA**

Course Level	Transfer Level		One Level Below Transfer		Two Levels Below Transfer	
	Pearson *r*	Corr. *r* *	Pearson *r*	Corr. *r* *	Pearson *r*	Corr. *r* *
Placement Test						
APS English	.25	.36	.30	.41	.29	.41
N	*626*	*626*	*895*	*895*	*315*	*315*
APS Reading	.18	.27	.20	.30	.11	.17
N	*626*	*626*	*895*	*895*	*315*	*315*
MTDP Math	.32	.42	**	**	**	**
N	*366*	*366*				

* Corrected for restriction of range

** A=4, B=3, C=2, D=1, F=0, W, Drop, NCr INC=Missing

Findings

The data in Table 2 indicate that the correlations between the placement tests and the dependent variable of final grade received in respective English or mathematics course was weak. Because students included in the study were already pre-sorted using an existing placement system, thus reducing the variability in the different levels of coursework, a correction for restriction of range was applied to the data. Although this improved the correlations somewhat, in the case of the reading test, the resulting correlations were still below the .35 mandated by the state chancellors office for adequate test validation evidence.

The low correlations obtained in response to research question one led directly to the need for the identification of other variables to construct a model that would more accurately and fairly place students. This led to research question two of this study.

However, these weak correlations also generated other questions that merited investigation. Why were the correlations so low? Was there a dramatic mismatch between what the assessment tests measured, and what was needed to succeed in class? As noted in the review section of this study, a review of prior research conducted in different educational and institutional settings suggested that low correlation coefficients between the variables of interest would occur. Additional review of the test validation literature suggested three primary reasons for the observed low correlation coefficient between placement score and final grade. These constraints may severely limit the strength of the correlation coefficient and include:

1. Restriction of range in the criterion variable of final grade,
2. Restriction of the range on the predictor due to use of an existing cut score to place students, and,
3. The instability of the criterion measure of final grade due to instructor grading variability.

Investigation of these constraints in the data base was conducted in the analysis for research question two. There was little to be done about restriction of range in the criterion variable of final grade. The database did not have another indicator of student success that was continuous and hence useful for multiple linear regression. Regarding constraint number two, analyzing data in an existing educational institution did not enable the suspension of the assessment and placement system to allow students to enroll in whatever level of English or mathematics they wished. With an existing placement system in operation using retrospective data, the truncation of the distributions in different levels of courses was unavoidable. Although it is possible to apply a correction for restriction of range as was done in Table 2, this is of little use in validating the use of a particular score in determining placement.

Instructor Grading Variation

It was possible to investigate the instability of criterion variable due to instructor grading variability. The SDCCD database contained instructor identification codes that could be used to

test for significant differences in mean GPA in different classes. Coded as dummy variables, the instructor ID could be entered into the regression model built to explain and better predict final course grade. This is addressed further on in this study.

RESEARCH QUESTION TWO: EXPLAINING VARIANCE IN FINAL COURSE GRADE

To determine what variables to include in the regression equation, several crosstabulations were conducted to examine the association between several of the student input, affective, commitment, environmental, educational background variables, with final grade in English or mathematics class as the dependent variable. For this analysis, final grade was recoded into "Success," (A,B, C, or CR) and "Non-Success" (D, F, NCR). If a significant relationship were found based on a chi-square or phi statistic, this variable was selected for inclusion in the multiple regression model. Some examples of these associations are illustrated below.

Input Variables and Final Grade. Figure 1 provides an example of a crosstabulation of student race/ethnicity and final grade in transfer level English courses. This crosstabulation suggests an association between certain ethnic/racial groupings and success in transfer level English. For example, Asian and White students appear to have higher rates of success in these courses. This observation was confirmed by the chi-square statistic which was significant at the .05 level.

Educational Background and Course Success. Figure 2 provides suggests the value of high school GPA for inclusion in the regression model. As self-reported high school GPA goes up, so does the rate of success. These observed associations with also significant at the .05 level.

Figure 1

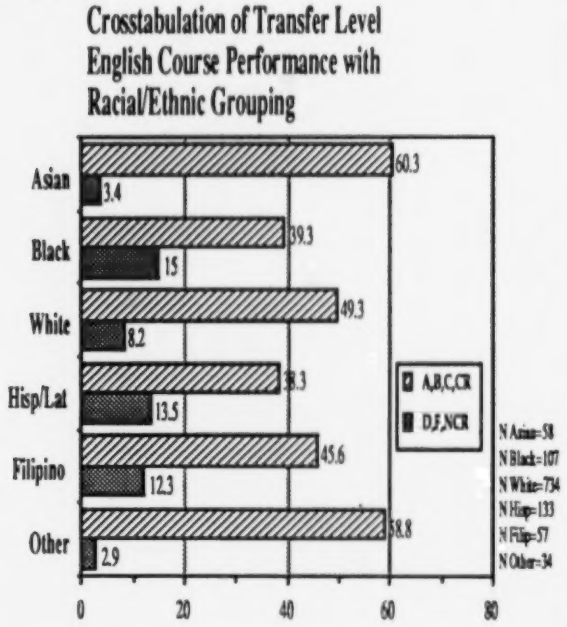
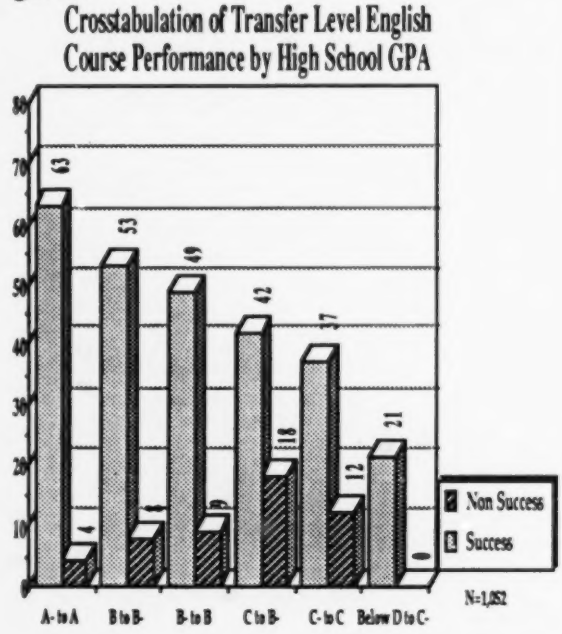


Figure 2



Commitment Variables and Course Success. Prior research has suggested a strong relation between student commitment to college attendance, involvement, the influence of peers and "significant others" on the student and success in college (Astin 1991; Astin, 1985; Cohen & Friedlander, 1989). Examples are provided below of the associations between selected "commitment to college" and peer and significant other variables with the criterion variable of course success. These crosstabulations also suggest that these variables are positively related to course success and inclusion in the linear regression model. Chi-square analysis also revealed that these variables were significantly related at the .05 level.

Figure 3

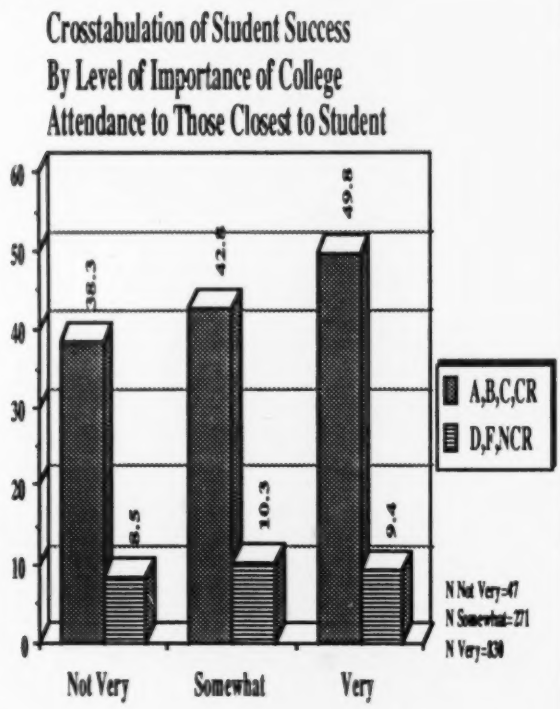
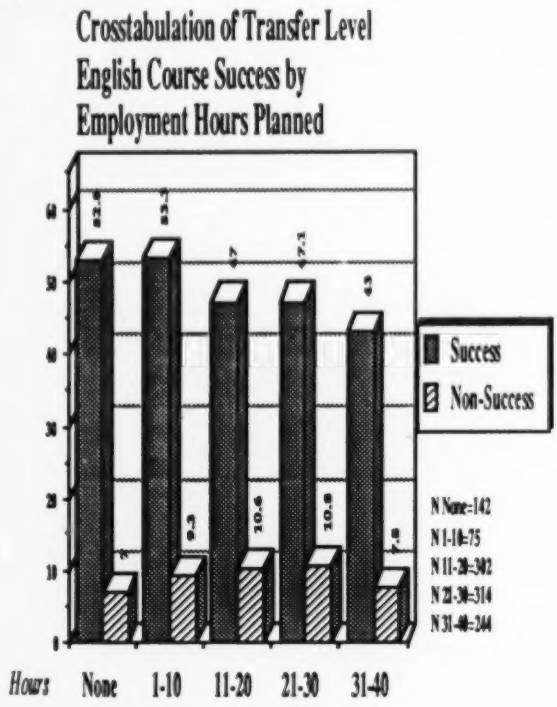


Figure 4



Grading Variability

As noted above, the instability of the criterion variable of final grade may severely limit the strength of the observed correlation coefficients between test scores and final grades. Moreover, if substantial grading variation did exist, any model attempting to explain or predict student success in courses, the need for pre-requisites, or college for that matter, would be unstable and fail to explain or predict very well. To determine differences between instructors, a one-way ANOVA for independent data was performed with student grades. An example of this analysis for transfer level English courses during the Fall, 1990 semester is shown in Table 3.

TABLE 3
TRANSFER LEVEL ENGLISH
STUDENT PERFORMANCE DATA
BY INSTRUCTOR

Instructor Code	Mean GPA	Std. Dev	Cases
01	2.96	.47	32
02	2.83	.84	29
03	2.92	1.01	26
67	3.14	.81	27
46	2.31	1.28	32
49	2.46	1.50	28
84	3.00	.92	22
74	3.98	.14	51
17	2.84	1.09	33
37	3.20	1.09	30
64	1.73	1.26	34
86	2.91	.93	49
56	2.23	.83	13
20	2.41	1.18	41
38	2.26	1.38	30
45	2.33	1.15	21
68	3.16	.69	36
83	2.84	1.06	19
56	2.62	.62	29
05	2.21	.95	23
13	2.95	.74	26
44	2.92	.74	26
61	2.51	1.22	27
Total	2.78	1.00	680

Summary One-Way Independent ANOVA on GPA by Instructors

Source	Sum of Squares	df	MS	F	P
Between Instructors	171.23	22	7.78	7.78	.0000
Within	657.11	657	1.00		

Eta= .4547 **Eta Squared=** .2067

Depending on the observed *eta* statistic, the instructor code would be included in the final model for explaining variance in course grade. In this example, approximately 20% of the

variance in final grade for transfer level English is explained by the instructor code. Similar results were found for courses one level below transfer level (Table 4) and two levels below transfer (Table 5).

TABLE 4
STUDENT PERFORMANCE DATA
ONE LEVEL BELOW TRANSFER ENGLISH
BY INSTRUCTOR

Instructor Code	Mean GPA	Std. Dev	Cases
90	2.53	1.02	19
68	2.29	.87	48
57	2.11	.68	56
35	2.23	.76	26
40	2.41	.94	17
03	1.65	1.02	26
83	2.73	.82	64
78	2.29	.81	28
87	2.73	.74	30
17	2.45	.68	31
50	3.13	.86	55
88	2.40	.89	30
39	2.84	.87	43
Total	2.50	.83	473

Summary One-Way Independent ANOVA on GPA by Instructors

Source	Sum of Squares	df	MS	F	P
Between Instructors	64.70	12	5.39	7.76	.0000
Within	319.54	460	.69		

Eta= .4104 **Eta Squared =** .1684

TABLE 5
SAN DIEGO COMMUNITY COLLEGE DISTRICT
ENGLISH 50 STUDENT PERFORMANCE DATA
BY INSTRUCTOR

Instructor Code	Mean GPA	Std. Dev	Cases
02	2.39	.94	23
79	3.22	.70	27
16	2.33	.87	9
89	2.47	1.12	17
35	3.00	1.00	15
17	2.87	.99	15
98	2.18	1.22	22
12	2.57	.81	21
Total	2.65	.96	149

Summary One-Way Independent ANOVA on GPA by Instructors

Source	Sum of Squares	df	MS	F	P
Between Instructors	19.32	7	2.76	2.98	.0060
Within	130.53	141	.93		

Eta= .3591 Eta Squared = .1289

These results suggested the presence of substantial grading variability across the curriculum from approximately 13% for two levels below transfer level, to almost 21% for transfer level English courses. The results for mathematics, while not shown here, also suggested substantial grading variation among instructors (from 6% to 15%). However, because other factors might account for this such as more poorly prepared students, or the time of day course is taught, it was decided that these other variables should be controlled for to test the extent of grading variation. Although the original intent of the model was to apply it to the college setting the strengthen the placement process, the foregoing analyses suggested that

instructor grading variation should be included in the model predicting final grade. To address these questions and to answer question two of this study, a multiple linear regression was run using the procedures outlined below.

Method

Ordinal level data and continuous data were left uncoded except in the case of variables found in the crosstabs not to exhibit a linear relationship with the criterion variable. In this case the data were recoded into dummy variables. Instructor and other nominal level variables such as environmental, were also recoded into dummy variables using methods recommended by Astin (1991) for use in multiple regression procedures in SPSS. Data were blocked and entered in the following way:

1. The first set of variables entered were the scores on the reading and English placement tests. In the case of mathematics, the MTDP test was the first variable entered. These were "forced" into the equation to function as proxy "pre-tests" for final grade.
2. Student input variables were next. These data included variables such as race, sex, age, income, native language, and disability. These were blocked and allowed to enter based on a stepwise routine.
3. Environmental variables were next allowed to be entered using stepwise methods.
4. Educational variables were the next block allowed to enter next using a stepwise procedure. These data included high school GPA, grade in last English course, high school last attended, years out of school, educational goal, etc.
5. The next set of variables were the "commitment' variables. These included affective data such as importance of college, employment hours planned, units planned, matriculation status, certainty of major, importance of college to those closest to student, etc.
6. Finally, the instructor dummy variables were entered as a block.

Table 6 summarizes how the model was constructed and the order in which the variables were blocked and entered or allowed to enter depending on their respective explanatory power.

TABLE 6

MODELING OF VARIABLES USED TO EXPLAIN VARIANCE IN COURSE GRADES BY GROUPED BY BLOCKS AND ORDER OF ENTRY INTO REGRESSION EQUATION

English Courses	Mathematics Courses
Input and Pre-Test	
APS Placement Tests (ETS)	MTDP Tests (CSU/UC)
Reading	Algebra Readiness
English	
Sex, Age, Race, Disability, Native Language, Income,	Sex, Age, Race, Disability, Native Language, Income,
Educational Background	
High School GPA, Highest Level English Course in High School	High School GPA, Highest Level Math Class in High School
Grade Received in Last English Class, HS Education, High School Attended, Interaction HS English GPA *and* Years of HS English Years Out of High School Veteran Status	Grade Received in Last Math Class, HS Education, High School Attended, Interaction of HS Math GPA *and* Highest Math Level Years Out of High School Veteran Status
Environment	
College of Attendance, Matriculation Status, Day or Evening Attendance	College of Attendance, Matriculation Status, Day or Evening Attendance
Commitment	
Employment Hours Planned	Employment Hours Planned
Importance of College to: Those Closest to Student *and* Student	Importance of College to: Those Closest to Student *and* Student
Educational Goal (transfer, personal, voc.)	Educational Goal (transfer, personal, voc.)
How Definite Choice of Major	How Definite Choice of Major
Units Planned this and Next Term	Units Planned this and Next Term
Grading Variability	***Grading Variability***
Instructor ID	Instructor ID

Final grade (GPA) was the criterion measure. This was left as a continuous variable on a five point scale (A=4, B=3, C=2, D=1, and F=0) with W's, Drops, Incompletes, and No Credits not included.

BEST COPY AVAILABLE

Explaining Variance in Grades for English: Findings

One hundred twenty variables were entered for transfer level English courses; 105 for courses one-level below college, and 85 for two levels below college. The change in R square was noted at each step to determine the amount of variance accounted for by the four blocks of independent variables. The summary table for the transfer level English course regression analysis is reproduced below in Table 6.

TABLE 6

SUMMARY TABLE OF REGRESSION OF FINAL GRADE ON INDEPENDENT VARIABLES FOR TRANSFER LEVEL ENGLISH COURSES

Step	Mult R	R^2	F	Sig. F	Variable:	Beta in
1					English Test	.2633
2	.2829	.0800	18.00	.000	Reading Test	.1150
3	.3135	.0983	15.00	.000	Income: $21-$26999	.1359
4	.3394	.1152	13.40	.000	Veteran	.1328
5	.3548	.1259	11.84	.000	Race:Asian	.1049
6	.3682	.1356	10.71	.000	Race:White	.1072
7	.4077	.1662	11.64	.000	HSGPA:2.0-2.5	-.1797
8	.4189	.1754	10.85	.000	HSGPA:1.5-2.0	-.0969
9	.4281	.1832	10.14	.000	Years out school>10	.0894
10	.4392	.1929	9.70	.000	Employment hours=0	.1024
11-57	***.6178***	***.3817***	***3.88***	***.000***	***Instructor ID***	****

(N=417)

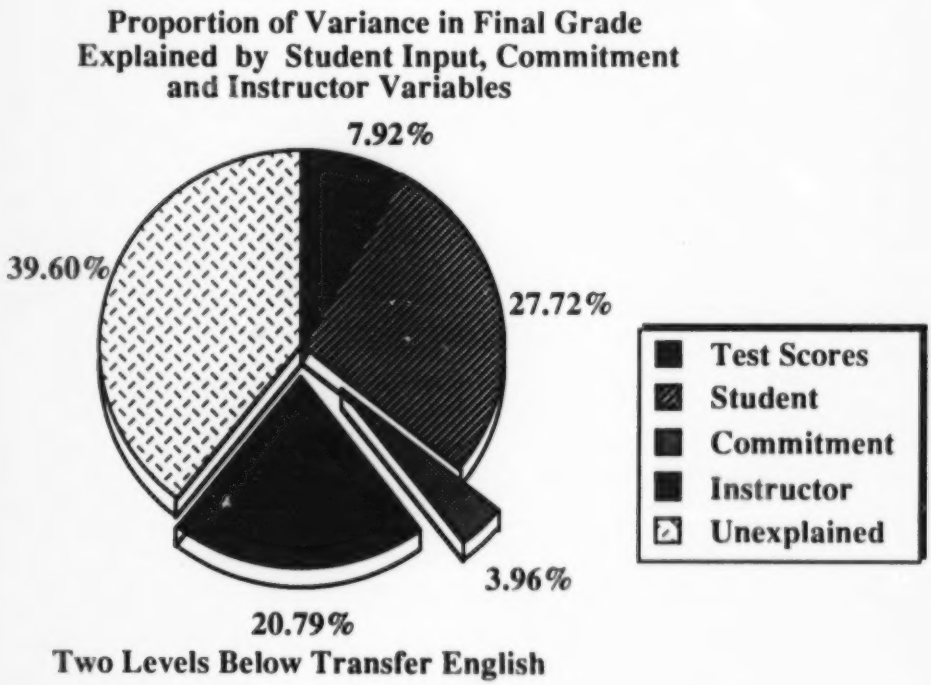
After controlling for several input and educational background variables, and attempting to control for several which did not enter the equation, this analysis revealed that approximately 19% of the variance in final grade in transfer level English courses is accounted for by the instructor codes. This is evidenced by the significant change ($p<.001$) in R square at the step

where instructor codes are entered. This is further evidence of substantial grading variation and may help to account for the low correlations observed in the analysis for question one.

One and Two Levels Below Transfer Level

Similar analyses were conducted for the other two levels of English courses. The results for one level below English showed that instructor grading variability accounted for approximately 14% (final R square of ..3775) of the variance in final grade, while two levels below transfer level English, the change in R square resulting from the entry of instructor IDs into the equation was approximately .217, (final R square of .599) or almost 22% of the variance in final grade was determined by the instructor after controlling for several key input and environmental variables. The proportion of variance in the criterion variable for two levels below transfer English is illustrated in Figure 5.

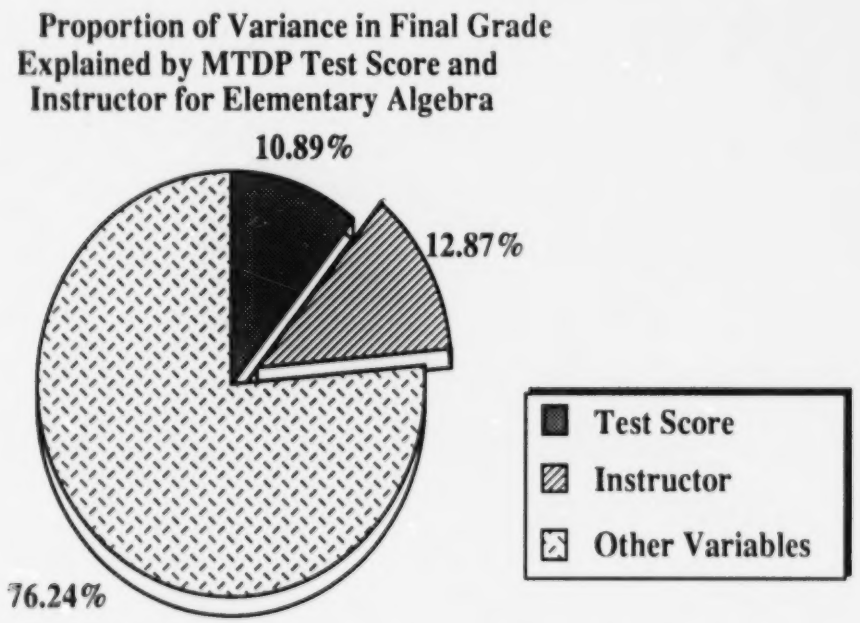
Figure 5



Mathematics

The analysis for mathematics took a somewhat different approach. Here the research focused solely on the proportion of variance explained in final grade when only the mathematics placement test and instructor grading practices were included in the regression. These results are illustrated below for Elementary Algebra classes in Figure 6.

Figure 6



This analysis suggests that in the case of Elementary Algebra, the placement tests explains less of the variance in final grade than the variability accounted for by the instructor. Here the proportion in variance accounted or explained by the instructor is approximately 13%. This finding obtained even though mathematics is often considered more "objective" in terms of course content and evaluation of student learning for various competencies in mathematics.

Summary and Implications

The difficulty in using assessment tests in predicting final grades seems evident. This should not be surprising given the often noted difficulty of predicting who will succeed in training, on the job, or in college. Since their first widespread application as a screening and sorting tool in the military services during World War I, standardized tests have been increasingly used as a predictor of individual success, often with mixed results, both in the civilian and military sectors. Knowledge of other relevant variables pertaining to individual aptitude and motivation appear to be essential to making informed decisions about who to admit, and where best to place them.

However, even with assessment of other relevant and related information about student aptitude and commitment, it appears that any model that seeks to use final course grade as the criterion variable for determining the validity of a placement rule, the need for a pre-requisite, or the prediction of student success will likely fail to account for a significant source of variation if instructor grading variability is not included. This finding has implications not only for state test validation policy, but for instructional issues across the curriculum. This is particularly true in the community colleges where the teaching and learning mission is first and foremost.

One need not be on a campus very long before realizing that all instructors who teach the same course do not grade alike. Students who are around a while also seem to learn this. For example, in the preliminary analysis of the correlation between test score and final grade, negative correlations were found for some courses. In discussions with colleagues, they noted that this suggested that students who had low ability in English or reading were taking courses from instructors known to be easy graders. This may in part help to explain the occurrence of these negative correlations.

Although teaching styles and grading policies and practices have much to do with student success rates, we do not enter instructor variables into the equation for predicting success because student selection of a specific section and instructor usually comes after participation in pre-enrollment matriculation services. Use of instructor codes in test validation research is also

politically sensitive. Once the threshold of the classroom is breached it is argued that academic freedom will be threatened. It is here, behind the classroom door, that state and community concerns with test validation and instructional accountability may confront closely guarded academic freedom. However it appears that in the interest of student equity and access, these critical instructional issues should be discussed and used in to inform departments and faculty of this important issue and how it impacts the student.

The issue of grading variability should not be ignored. With the primary mission of providing instruction, community college faculty, particularly those in the English or Mathematics departments might use these results to critically examine the goals and desired outcomes of various course levels. This is particularly important when one considers that English or Mathematics prerequisites are required for many transfer and sophomore level courses in the two year colleges. Articulation agreements to facilitate student transfer also hinge on the perception that there are standards and objectives of instructional quality. The findings suggest that individual departments should carefully review course objectives and evaluation processes. This would probably better serve the student, and improve the processes for internal evaluation of the instructional programs. Grading variability not only makes matriculation evaluation difficult, it also raises equity concerns for students.

REFERENCES

Alkin, M.C. & Freeman M. (1991). *Matriculation evaluation: Summary report* 1990-1991. Northridge CA: Educational Evaluation Associates.

Asher, J., & Sciarrino, J. (1974). Realistic work sample tests. Personnel Psychology, 27, 519-533.

Astin, Alexander, W. (1991). **Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education**. New York, Macmillan

Astin, Alexander, W. (1985). **Achieving educational excellence: A critical assessment of priorities and practices in higher education**. San Francisco, CA: Jossey-Bass.

Astin, Alexander, W. (1985a). *Excellence and equity: Achievable goals for American education*. Phi Kappa Phi Journal 64(2), pp 24-29.

California State Assembly, (1986). *Assembly Bill No. 3; Ch. 1467 Legislative Council's Digest of the Seymour-Campbell Matriculation Act*. Sacramento, CA. California State Publications

Cohen, A.M., and F. Brawer (1988)*The American community college*. San Francisco: Jossey-Bass

Cohen, E.& J. Friedlander (1989). *Approaches to predicting student success: Findings and recommendations from a study of California Community Colleges*. Santa Barbara, CA. Santa Barbara City College.

Cronbach, Lee. J. & G. C. Gleser (1965). **Psychological tests and personnel decisions**. Urbana, Illinois: University of Illinois Press.

Gillespie, Maggie (1993, Feb.) Placement testing in community colleges: A Response to Hughes and Nelson. Community College Review; 20 (4), pp 59-69.

Hughes, Ronald E. & C.H. Nelson (1991). Placement scores and placement practices: An empirical analysis. Community College Review 19(1), pp 42.46.

State Chancellors Office (1992). *Standards, policies and procedures for the evaluation of assessment instruments used in the California community colleges*. Sacramento, CA: Matriculation Assessment Work Group, California Community Colleges and Center for Educational Testing and Evaluation, University of Kansas.

State Chancellors Office (1992, April). Matriculation Regulations: Sections 51024-58508. Sacramento, CA.

Sticht, Thomas G. (ed.) (1975). **Reading for working: A functional literacy anthology**. Alexandria VA: Human Resources Research Organization.

Sticht, Thomas G., Armstrong, W.B., Hickey, D.T., and Caylor, J.S. (1986). **Cast-off Youth: Policy and training methods from the military experience**. Westport, CT: Praeger.

The American College Testing Program (ACT) and the American Association of Community and Junior Colleges (AACJC), (1985). *Status of testing practices at two-year postsecondary institutions*. Iowa City, Iowa: ACT Publications.

The College Board (1984) Assessment and placement services for the community colleges: Using and interpreting scores. Princeton NJ: College Entrance Examination Board.

The National Commission on Testing and Public Policy, (1990). *From gatekeeper to gateway: Transforming testing in America*. Chestnut Hill, MA: Boston College